5-2-2013

Reliability-aware multi-segmented bus architecture for photonic networks-on-chip

Patrick Sieber

Reliability-Aware Multi-Segmented Bus Architecture for Photonic Networks-on-Chip

by

Patrick Sieber

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Engineering

Supervised by

Dr. Amlan Ganguly
Department of Computer Engineering
Kate Gleason College of Engineering
Rochester Institute of Technology
Rochester, NY

Approval Date: May 2nd, 2013

Approved By:

_____

Dr. Amlan Ganguly
Primary Advisor – R.I.T. Dept. of Computer Engineering

_____

Dr. Kenneth Hsu
Secondary Advisor – R.I.T. Dept. of Computer Engineering

_____

Dr. Sonia Lopez-Alarcon
Secondary Advisor – R.I.T. Dept. of Computer Engineering

Abstract

Network-on-chip (NoC) has emerged as an enabling platform for connecting hundreds of cores on a single chip, allowing for a structured, scalable system when compared to traditional on-chip buses. However, the multi-hop wireline paths in traditional NoCs result in high latency and energy dissipation causing an overall degradation in performance, especially for increasing system size. To alleviate this problem a few radically different interconnect technologies are envisioned. One such method of interconnecting different cores in NoCs is photonic interconnects. Photonic NoCs are on-chip communications networks in which information is transmitted in the form of optical signals. Photonic interconnection is one of the leading examples of emerging technology for on-chip interconnects.

Existing innovative photonic NoC architectures have improved performance and reduced energy dissipation. Most architectures use Wavelength Division Multiplexing (WDM) on the photonic waveguides to increase the data bandwidth. However they have issues relating to reliability, such as waveguide losses and adjacent channel crosstalk. These phenomena could have a crippling effect on a system, and most current architectures do not address these effects. A newly proposed topology, known as the Multiple-Segmented Bus topology, or MSB, has shown promise for solving, or at least reducing, many of the problems plaguing the design of photonic networks using a modification of a folded torus to transmit different wavelength signals simultaneously. The MSB segments the waveguides into smaller parts to limit the waveguide losses. The formal performance evaluation

of this proposed architecture has not been completed. This thesis will analyze the performance of such a network when implemented as a NoC in terms of data bandwidth, energy dissipation, latency, and reliability. By analyzing and comparing performance, energy dissipations, and reliability, the MSB-based photonic NoC (MSB-PNoC) can be compared to other state-of-the-art photonic NoCs to determine the feasibility of this topology for future network-on-chip designs.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

With increasingly difficult and complex design challenges, the need for continually more and more powerful processing is a very real issue. However, a simple increase in the number of transistors and frequency of clock rates is proving to be increasingly difficult, becoming altogether impractical in recent years. As frequency scales upwards, so does power, due to higher switching activity and higher power density, which opens up an entirely different set of problems. With power increases come battery life issues, excessive heat, and many other prohibitive issues that prevent frequency increase from being a practical way to increase performance. [1]

1.1. Introduction of Multi-Core Systems

One accepted course of action to address power concerns has been a shift towards multi-core systems. Instead of running one core at a higher speed, several lower-speed cores run simultaneously, dividing up the workload and parallelizing the execution. This allows frequencies to remain low, eliminating many of the problems of single core systems. However, this introduces the new problem of how to connect the multiple cores. With ever-increasing numbers of cores, the design of the interconnections becomes critical.

1.2. Network-on-Chip as a means of interconnecting Multi-Core System-on-Chip

Systems-on-chip are distributed systems on a single silicon substrate. This allows for globally asynchronous and locally synchronous setups, using many different clocks, which eliminates the probability of excessive clock skew when a single clock source is used across a large system [2]. Interconnection of hundreds of cores in current and future multicore chips will be enabled by the Network-on-Chip paradigm. The concept itself comes from the "route packets, not wires" paradigm [3]. This allows for the separation of the data transport infrastructure from the functionality hardware. This decoupling creates a dedicated infrastructure for the communication of the system, allowing for a more modular design. Wireline connections on such systems, however, draw large amounts of power, and also exhibit large amounts of signal degradation, in addition to high latency. In fact, the International Technology Roadmap for Semiconductors even predicted that 80% of chip power would be because of the on-chip interconnects alone [13]. Clearly, this points to the fact that novel and revolutionary technology is necessary to circumvent the problem of power consumption in future generations of multicore chips.

1.3. Emerging Technology

Some of the methods used to alleviate many of these problems include 3-D integration, wireless and RF interconnects, and high-bandwidth and low-energy

photonic links. 3-D integration, for example, involves stacking multiple layers of circuitry. This results in more interconnections, as each core has another axis along which to link. The stacked cores allow for shorter interconnects overall, since cores have more immediate neighbors [4]. However, because of the higher core density due to the smaller 2-dimensional footprint, the heat and power densities are increased, making high temperatures a problem. Stacking of layers also opens up the possibility for manufacturing defects creating mismatches between the layers, making them incompatible with one another. Wireless on-chip networks use RF wireless interconnections to connect some or all cores. The most common usage of this technology is to connect distant cores, where wireline links would show the greatest performance penalty. By using carbon nanotube technology to create antennas, cores are shown to be able to communicate [5]. This solves the degradation problem of long wires, but introduces challenges in creating reliable wireless links, as well as dealing with wireless link failures. Of course, the system requires precision wireless transceiver hardware to be introduced as well.

1.4. Photonic NoCs

Another state-of-the-art technology being researched is photonic networks on chip (PNoC). This technology uses the high-bandwidth benefit of photonic links for high payload transfers. By using the low loss properties of optical waveguides to send information, higher bandwidth, lower latency, and lower power dissipation can be achieved compared to fully electronic NoCs. The waveguides also have low levels of loss, allowing data to be transmitted end-to-end without the need for repeating,

regenerating, or buffering, which is also a large improvement over electronic networks [1]. By using dense wave-division multiplexing (DWDM), single buses are able to transmit waves simultaneously at different frequencies. This allows for increased bandwidth when compared to the number of photonic links. Photonic networks also only need to have photonic switches turn on once per message, as opposed to once per bit like electronic network, which makes energy dissipation independent from bit rate, further decreasing the overall energy dissipation [6]. Photonics are particularly effective for global interconnects, allowing for easier scalability. As with any NoC, there are issues with signal degradation and crosstalk. To remedy these, there are several different interconnect configurations that attempt to alleviate the problems by changing the way cores are connected to one another. However, these architectures were designed to improve performance of the system, but reliability has not been taken into account sufficiently. As a result, many have issues with signal loss, especially across long links, as well as unpredictable latency and congestion issues. A reliability-aware Photonic NoC technology is the main focus of this research.

1.5. Thesis Contributions

In this thesis work it will be demonstrated that by using a proposed PNoC design known as the Multi-Segmented Bus (MSB), high data throughput and lower energy dissipation can be achieved while maintaining reliable data transfer. The following is a summary of contributions made in this research.

- **Proposed Architecture Model**
    - Architecture of the proposed PNoC
    - Design the MSB based PNoC for 64, 128, and 256-core systems, including core-to-core connections and routing paths.
- **Experimental results**
    - Performance evaluation of the proposed MSB based PNoC using a cycle-accurate simulator.
    - Obtain experimental results of the proposed MSB architecture, as well as other PNoC architectures in state-of-the-art literature for comparison, with respect to the following parameters:
        - Bandwidth
        - Packet energy dissipation
        - Bit-error-rate (BER) in data transmission
        - Scalability - Increasing system sizes
        - Non-uniform traffic patterns (Hotspot, transpose, FFT)

- **Publications**
    - Pradheep Khanna Kaliraj, Patrick Sieber, Amlan Ganguly**,** Ipshita Datta, Debasish Datta, “Performance Evaluation of Reliability Aware Photonic Network-on-Chip Architectures”, IGCC Workshop on Lighter than Green Reliable Multicore Architectures, International Green Computing Conference (IGCC), San Jose, 2012.

Chapter 2 Related Work

There are a variety of NoC architectures for photonic NoCs. Some of these include a 2-Dimensional Folded Torus (2DFT), Corona, and Clos.

2.1. 2-Dimensional Folded Torus

2DFT is one of the most commonly studied architectures for PNoCs because it has been physically realized. In 2DFT, each cluster contains a gateway switch (GS), an ejection switch (ES), an injection switch (IS), and a network switch (NS). These switches allow each cluster to send and receive packets, as well as route them to their appropriate destinations [7]. These switches use Microring Resonators (MRR) to direct light waves along different paths towards the intended destination. MRRs have a vital building block for photonic systems. The small size allows for low power operation and dense integration, and their wavelength selectivity allows for cascaded wavelength division multiplexing (WDM) [8]. They work by using a resonant frequency, and if the lightwave matches that frequency, the wave is pulled along the ring, allowing the signal to be routed along a different path. Otherwise, the wave continues through unchanged.

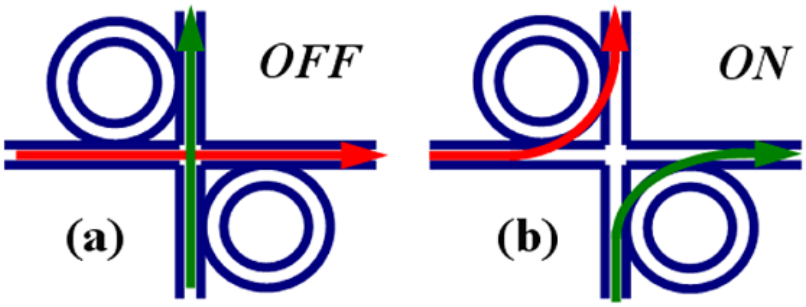


Figure 2-1: Inter-Segmented Router Behavior

The photonic paths are formed by a set of rings, or tori, which link either vertically or horizontally adjacent clusters.

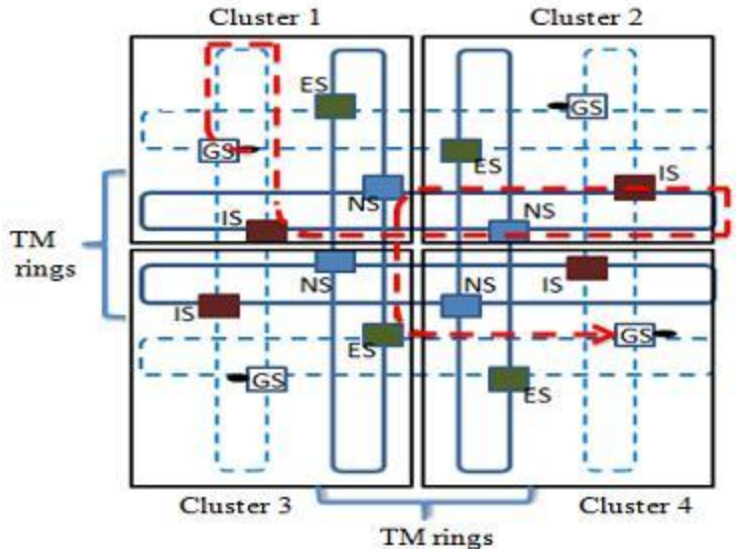


Figure 2-2 :2-Dimensional Folded Torus

The rings connect in the center of the system using a set of interleaved rings, allowing any cluster to communicate with any other cluster. However, the scope of the wavelength division multiplexing for this architecture is limited by the fact that each dedicated path must be tuned to a specific wavelength for the MRRs to work correctly, at a particular resonant frequency. To accommodate more wavelengths requires multiple torus rings as well as more MRRs, which increases the complexity of the system as well as the optical loss and crosstalk of the pathways. This has an adverse impact on the bit-error rate (BER) of the system [7].

2.2. Corona

The Corona architecture uses long waveguides running from a cluster through every other cluster back to itself, ending just before reconnecting to the initial end. The architecture needs a large number of waveguides, which get congested as the number of clusters increases. With more clusters also comes longer waveguides, which increases waveguide losses and crosstalk. This results in a decrease in BER as well [7].

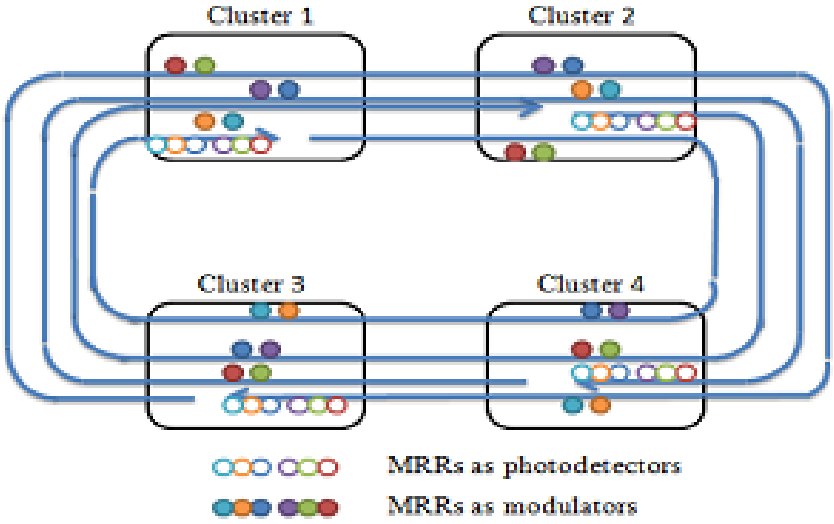


Figure 2-3: Corona Architecture

Corona clusters communicate using an optical crossbar, allowing a connection between every cluster [9]. Differently sized messages can simultaneously share the communication channels using WDM, provided they use different channels, in order to increase utilization. The clusters each have a designated channel for messages to share. All clusters can write to any channel, but only a single, specific cluster can read from any channel. Because of this, in order to

realize a fully-connected 64 x 64 crossbar must repeat the channel 64 times, with each cluster assigned as the single reader of one channel.

Each channel consists of 256 wavelengths, bundled into 4 waveguides. As light leaves the source, it passes through a splitter to distribute the wavelengths of light to the waveguide. The communication travels to each cluster in increasing order, looping around to the first cluster if need be. To send data to a cluster, the source cluster modulates the light on the channel read by the destination cluster [9].

2.3. Photonic Clos

Another popular architecture is Clos. A Clos system uses multiple stages of routers to create a larger non-blocking network. They are considered to be a midpoint between the crossbar topology, with its low diameter and high crossbar capacity, and the higher diameter mesh topology [10]. Clos routers are implemented electrically and the inter-router channels are implemented with photonics and are considered to enable flits to be transmitted in a single cycle.

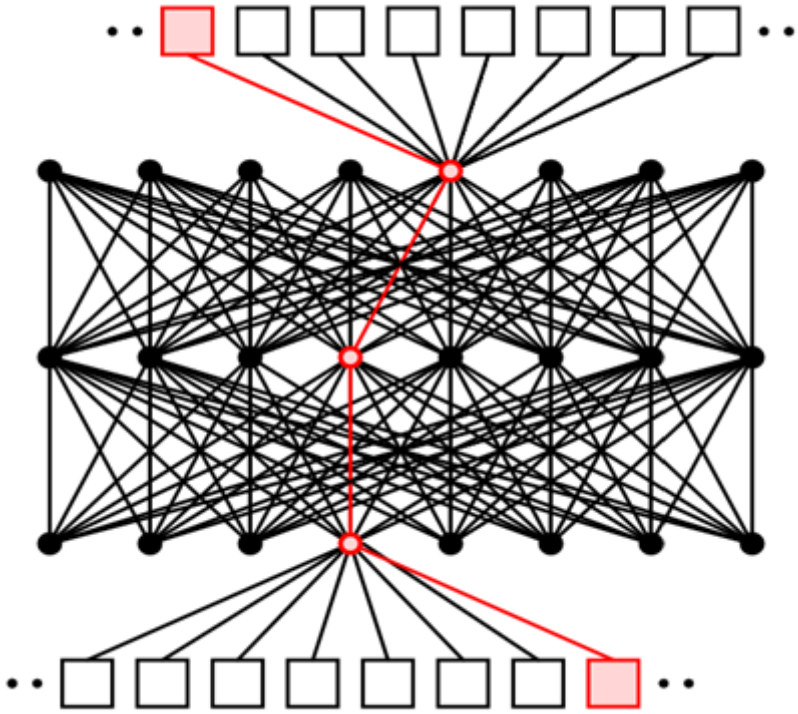


Figure 2-4: Clos Architecture

The architecture works by routing messages from the input through a series of middle routers to the output. Different routing algorithms can be used to choose which routers will be used in the path from source to destination. These are known as point-to-point channels. Another method of using Clos is by using photonic middle routers consisting of photonic crossbars. By routing using crossbars, one stage of conversion from electric signals to optical signals, then back to electrical signals, is removed. This can lower the dynamic power of routing, but usually results in an optical and thermal tuning power penalty. This tradeoff means that using electrical versus photonic routing is dependent on the specific system. The network also uses shorter waveguides and less rings along each waveguide than a full crossbar network. It is often seen as a viable replacement for crossbar networks because this causes a decrease in optical losses [10].

Another important feature of Clos networks is that they provide uniformly low latency and high bandwidth regardless of traffic pattern. This results in easier programming design, which can be an important factor in highly parallel systems.

In this work I propose the design of a scalable PNoC which has the best BER characteristics and evaluate its performance and compare with other PNoC architectures in literature.

Chapter 3 Reliability-Aware Photonic Architecture

The Multi-Segmented Bus based photonic NoC architecture is proposed as a way to take into account signal losses and crosstalk components to create a more reliable photonic architecture. This section will discuss the topology and routing of the MSB architecture, while the next chapter will discuss the reliability. The MSB uses the technology of the MRR for high bandwidth and low power designs. MRRs enable low-power operation and integration of hundreds of the device on-die because of their small footprint [1]. By taking advantage of wavelength selectivity, WDM can be used to increase the bandwidth of the photonic links. Figure 3-1 illustrates how MRRs are able to turn the light signals when switched on, allowing them to route the signals along multiple possible paths.

3.1. Topology

The MSB topology uses shorter buses, with each segment passing through a smaller number of clusters when compared to other configurations. Since longer segments result in a higher signal degradation over distance, having shorter segments limits the signal loss. To transmit over longer distances, the buses are linked using inter-segment routers (ISRs), which switch lightwaves from one bus to another. Turning these routers on and off uses MRRs to allow the path of the signal to be changed. These routers reduce the length of photonic connections traversed by a signal, reducing signal losses when compared to other existing PNoC architectures.

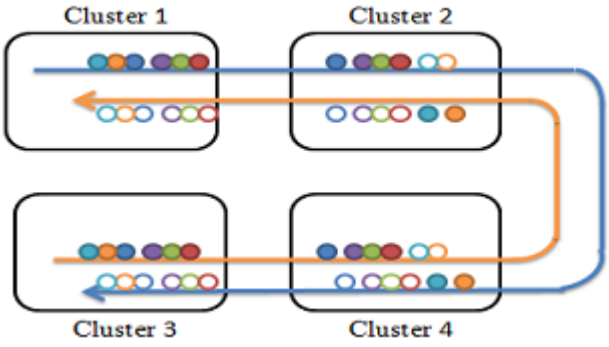


Figure 3-1: Multi-Segmented Bus Architecture

Figure 3-1 shows the basic construct of the MSB network without ISRs. In the MSB network, each link is segmented and arranged so that all of the segments, as well as the number of attached photonic devices, are the same as one another. This allows all segments to exhibit identical characteristics with respect to signal loss and noise. Each adjacent row of clusters (RC) is connected by a clockwise (CW) and counterclockwise (CCW) bus. This ensures that there is direct single-bus connectivity between RC pairs, shown generically as

$$RC(i[\mathrm{mod}-N]) - RC(i+1[\mathrm{mod}-N]) \qquad (1)$$

where N is the number of RC in a given NoC. Figure 3-2 shows a simple example of how clusters are connected when part of an adjacent RC. Vertically non-adjacent rows are connected by two MSB busses, which are joined together by an ISR. Through the use of these ISRs, there is a direct route from every cluster to every other cluster. A cluster can be composed of either a single core or multiple cores interconnected by electronic connections. This means that the system has full connectivity across all clusters, vastly simplifying the design process by eliminating the need to determine an "optimal" interconnection configuration. In order to prevent blocking along the bus lines, multiple parallel busses are needed between

rows of clusters. Figure 3-2 illustrates how the connections are formed between clusters, and shows the ISRs, indicated by the letter R, between MSBs. Any segment adjacent to one of the ISRs can use the router to transfer onto the other adjacent segment across that ISR.

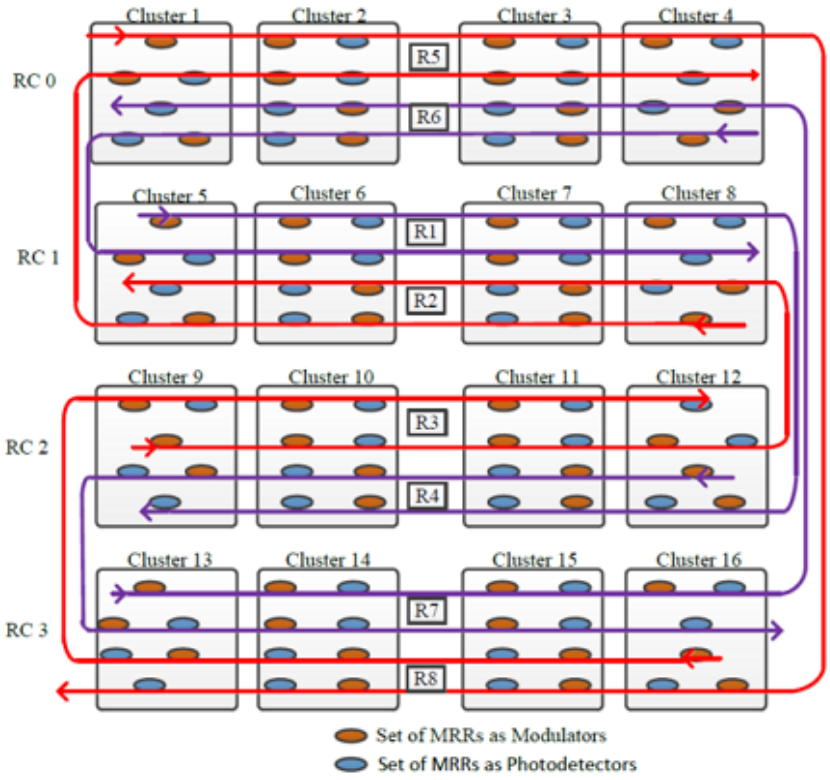


Figure 3-2: Larger, Connected Multi-Segmented Bus Architecture

One important aspect of this technique is that the size of the system can be scaled up quite easily from 16 clusters to 64, 128, or even 256 clusters by connecting groups of clusters using inter-group busses (IGB). In combining groups of 16 clusters like this, the top and bottom rows of each group are connected using the IGB, allowing a signal from any group to move to the IGB, then move to any other group. Figure 3-3 shows how four groups of 16 clusters are combined to form a 64-cluster system.

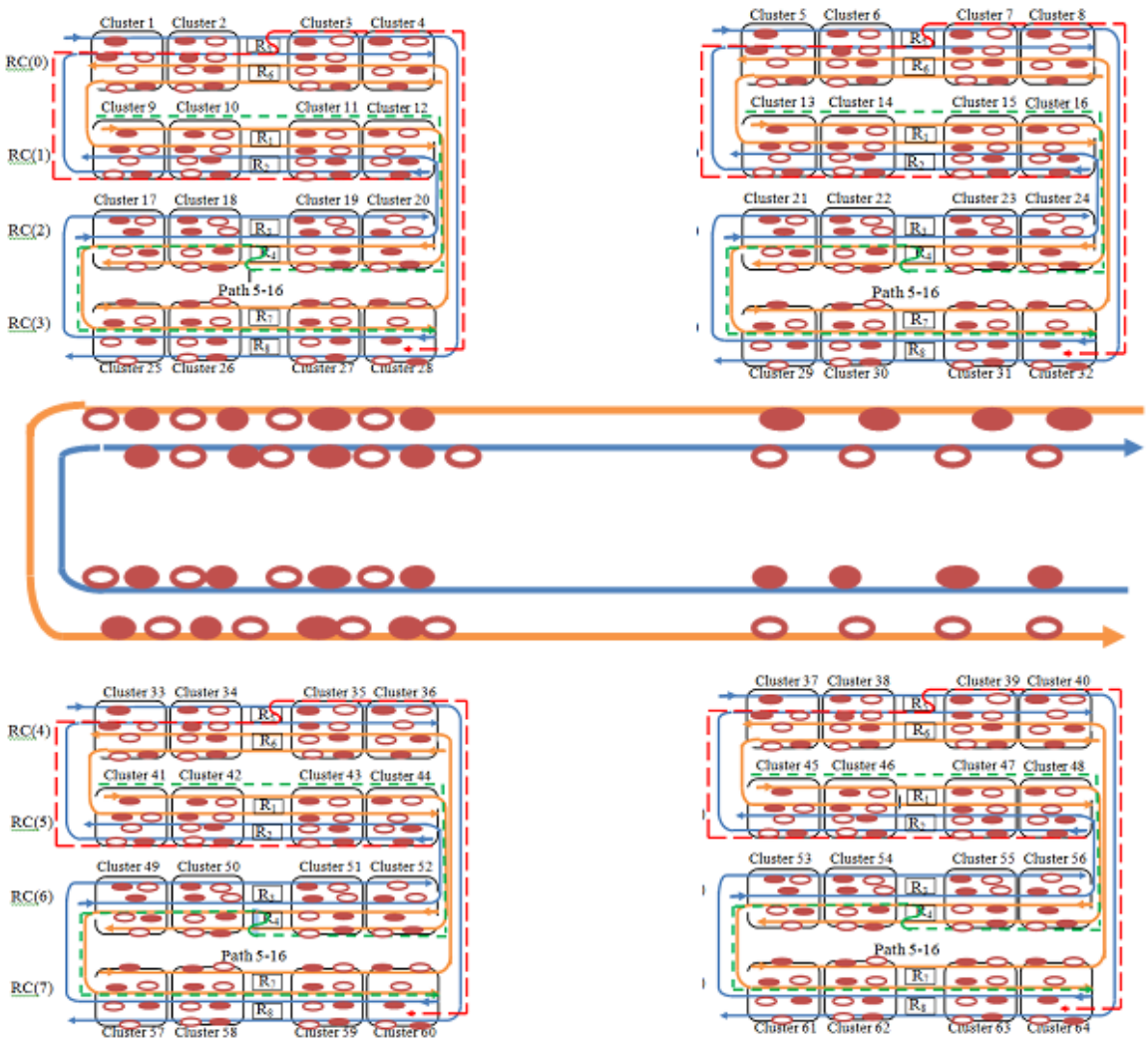


Figure 3-3 :64-Cluster Scaling, with IGB

3.2. Data Routing

Data is routed through the system using a packet switched routing protocol. Specifically, the system uses wormhole routing, which pipelines the network by dividing a message into packets, and further dividing those packets into flits. The flits are small enough to theoretically be transferred across any connection in a single cycle of the clock driving the NoC. In wormhole routing, the header flits have

the destination address, and the remaining flits making up the message simply follow the same path as the header. This allows the entire message to be moving through the links making up the path to its destination one cycle at a time.

If the source and destination clusters are part of the same 16-cluster group, all data is able to be transmitted solely on the MSBs. If the clusters are on vertically adjacent rows, the transfer is possible using a single MSB, otherwise a single MSB is not sufficient, and the ISRs are utilized to move the flits from one MSB to the next.

If the source and destination clusters are in different 16-cluster groups, the data will need multiple hops to reach the destination. In this case, flits travel from the source to the closest cluster connected to the IGB. The data is demodulated and converted back to the electrical domain so it can be moved into this cluster. It is then modulated back to the optical domain and moved to the IGB to be transmitted to the group containing the destination cluster. Upon reaching the destination group, the flits are again demodulated into the cluster connected to the IGB closest to the final destination. The data is then modulated once again onto the MSB within the cluster, and then transmitted to the final destination along the MSBs as in the other cases. As such, data travelling between different groups are transmitted over multi-hop paths and converted from the optical domain to electrical domain and vice versa. Clusters directly connected to the IGBs can transmit to the IGB in one hop using the IGB's modulators and demodulators and bypass the transfer from source MSB to IGB, saving a hop.

In a 256 core architecture, multiple IGBs exist to connect all of the clusters, and a transmission may require modulation and demodulation from one MSB to an

IGB to another MSB, increasing the number of total hops. Since the size of the flits is determined based on theoretically transmitting a flit across the segments in one clock cycle, traversing photonic links within one cluster will occur within one cycle, with an additional hop necessary to move the message from the MSB link to an IGB, and another additional hop to move from the IGB onto the photonic MSB link of another cluster. Consequently, for a signal to move from one cluster to another cluster in another group across the IGB and then from the cluster linked to the IGB to another cluster within that group, 3 cycles would be needed.

Chapter 4 Reliability Analysis

In this section, the Bit-Error Rate (BER) is evaluated for the MSB model being analyzed, as well as for other interconnect topologies. To model the reliability in data transfer, we consider two clusters, a distance apart, which have communication between two cores, one from each cluster. The lightwave received at the destination cluster in presence of crosstalk is expressed as:

$$E_R(t) = \sqrt{(2P_S(b_i))}\cos(2\pi f_s t + \theta_s + \phi_s(t)) + E_{XT}(t) \qquad (2)$$

The first term on the right hand side of (2) represents the signal component at the destination. $P_S(b_i)$ is the bit dependent received signal power, accounting for losses along the pathway, where $b_i \epsilon \{0,1\}$, f_s is the signal frequency, θ_s is the initial phase, and $\phi_s(t)$ is the phase noise of the signal component of the lightwave. Bit dependent received signal power is the power of the signal as it is received at the photodetector, accounting for all losses along the way. Phase noise describes fluctuations in the phase of the signal as it is transferred from source to destination. $E_{XT}(t)$ represents the accumulated crosstalk component given by

$$E_{XT}(t) = \sum_{j=1}^{W} \sqrt{(2P_{xj})}\cos(2\pi f_j t + \theta_j + \phi_j(t)) \qquad (3)$$

where W represents the number of crosstalk components, P_{xj} is the received power of the j-th crosstalk component, f_j is the frequency of the j-th crosstalk component, θ_j is the initial phase of the j-th crosstalk component, and $\phi_j(t)$ is the phase noise of the j-th crosstalk component. The photocurrent produced at the photodetector output is given by

$$i_p(t) = R_\lambda < E_R{}^2(t) > + i_{th}(t) + i_{sh}(t) \quad (4)$$

The first term on the right hand side of equation (4) defines the *square-and-average* operation of the photodetector on the received lightwave, with R_λ as the photodetector responsivity, the second term is the thermal noise of the receiver, and the third term represents the signal dependent shot noise. Thermal noise is electronic noise generated by thermal agitation of any conductor, and shot noise describes fluctuations in a photonic signal based on the locations of photons being independent of one another. The first term of the right hand side of equation (4) can be expressed as

$$R_\lambda < E_R{}^2(t) >= + i_s(t) + i_{sx}(t) + i_{xx}(t) \quad (5)$$

where $i_s(t)$ is the signal component of the photocurrent, $i_{xx}(t)$ and $i_{sx}(t)$ are the crosstalk-crosstalk and signal-crosstalk beat noise components. $i_s(t), i_{xx}(t)$, and $i_{sx}(t)$ are expressed as

$$i_s(t) = R_\lambda P_s(b_i) \quad (6)$$

$$i_{xx}(t) = R_\lambda [\sum_{j=1}^{W} P_{xj} + \sum_{j=1}^{W} \sum_{k=1}^{W} \sqrt{P_{xj} P_{xk}} \cos(\omega_{jk} t + \phi_j(t) - \phi_k(t) + \theta_j - \theta_k)] \quad (7)$$

$$i_{sx}(t) = 2R_\lambda \sum_{j=1}^{W} \sqrt{P_S(b_i) P_{xj}} \cos(\omega_{js} t + \phi_j(t) - \phi_s(t) + \theta_s - \theta_j) \quad (8)$$

where $\omega_{js} = \omega_j - \omega_s$ and $\omega_{jk} = \omega_j - \omega_k$ represent the respective beat-noise frequencies.

The combined electrical noise (shot noise, thermal noise, and signal-crosstalk beat noise (crosstalk-crosstalk beat noise is ignored here because it is relatively insignificant compared to the other values)) after photodetection is modeled as a zero-mean Gaussian random process with the variance expressed as

$$\sigma_{bi}^{\ 2} = \sigma_{sxi}^{\ 2} + \sigma_{shi}^{\ 2} + \sigma_{th}^{\ 2} \qquad (9)$$

where $\sigma_{th}^{\ 2}$ is the thermal noise variance with R as the input impedance, B_e as the noise equivalent bandwidth of the optical receiver, used to quantify leakage within the circuit, k as Boltzmann's constant, T as receiver temperature, and σ_{shi} represents the shot noise variance, given by

$$\sigma_{th}^{\ 2} = (4kTB_E)/R \qquad (10)$$

$$\sigma_{shi}^{\ 2} = 2q[R_\lambda P_s(b_i) + R_\lambda \sum_{j=1}^{W} P_{xj}]B_e \qquad (11)$$

The worst-case signal-crosstalk beat noise variance $\sigma_{sxi}^{\ 2}$ is given by

$$\sigma_{sxi}^{\ 2} = R_\lambda^{\ 2} \sum_{j=1}^{W} P_{xj} P_s(b_i) \qquad (12)$$

The receiver bit-error rate (BER) can be evaluated as

$$BER = P(1)P(0/1) + P(0)P(1/0) \qquad (13)$$

where P(0)and P(1) are the transmission probabilities of '0' and '1', and P(1/0) and P(0/1) are the respective conditional error probabilities. Under the Gaussian assumption for the probability density functions, the BER can be expressed as

$$BER = 0.5erfc(Q/\sqrt{2}) \qquad (14)$$

where $Q = R_\lambda\,[\mathrm{P_S}(1) - \mathrm{P_S}(0)]/(\sigma_1 + \sigma_0)$, *erfc* is the complementary error function, and the noise variances for the bits $\{b_i\}$ are given by

$$\sigma_{bi}^{\ 2} = R_{\lambda}^{\ 2} \sum_{j=1}^{W} P_s(b_i) P_{xj} + 2q[R_{\lambda} P_s(b_i) + R_{\lambda} \sum_{j=1}^{W} P_{xj}] B_e + (4kTB_e)/R$$
for $b_i \in \{0,1\}$ (15)

This BER evaluation method is adapted for all the PNoC architectures considered here while accounting for all the components of signal loss and interference.

Chapter 5 Experimental Results

In this section, the performance of the MSB-PNoC is evaluated and compared to a mesh architecture. Mesh was used as a main comparison because mesh interconnects are the main technology currently in use in physically creating this type of network. For some metrics, other photonic architectures were compared as well. In order to obtain results for the different architectures, a cycle-accurate simulator was used to model the behavior of an MSB system, as well as several other architectures for comparison. The main methods of comparison for the results are the peak bandwidth and packet energy dissipation. Peak sustainable bandwidth is the maximum rate at which the NoC is able to route data successfully. Packet energy is the average energy dissipated in transferring a data packet from source to destination. This analysis looked only at the energy dissipated in transferring from cluster to cluster, and ignored any energy dissipation within the clusters, in order to focus only on the contribution of the MSB architecture.

In the experiments, each cluster was considered to consist of a core and its associated switch. The switch architecture, as used in [11], has three stages: input arbitration, routing, and output arbitration. A cycle-accurate simulator uses this switch layout, with each switch is capable of modulating and demodulating data in order to transmit over the photonic links attached to its port. Converting data between the electrical and optical domains takes one clock cycle [9]. The port on each switch has 4 virtual channels containing a buffer with a depth of 2 flits. The cores are modeled at tiles in a 20mmX20mm die. The simulator monitors the flits' progression, tracking how many

reach the correct destination and how many are dropped. The simulations were all run over several thousand iterations to reach more stable results.

5.1. Performance-Reliability Trade-off

It has been shown in [7] that the BER of photonic links increases as bandwidth increases because of interference from adjacent frequency channels on the same bus which enable WDM. The BER model described in Chapter 4 can be used to calculate the BER in data transfer for photonic architectures. Figure 5-1 shows a comparison of calculated BER when using a 16-cluster system size with 20Gbps bandwidth links as a function of launched power, using a 20mmx20mm die for different PNoC architectures.

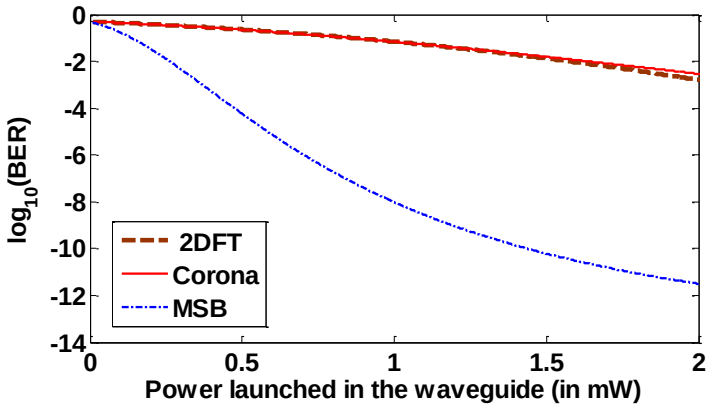
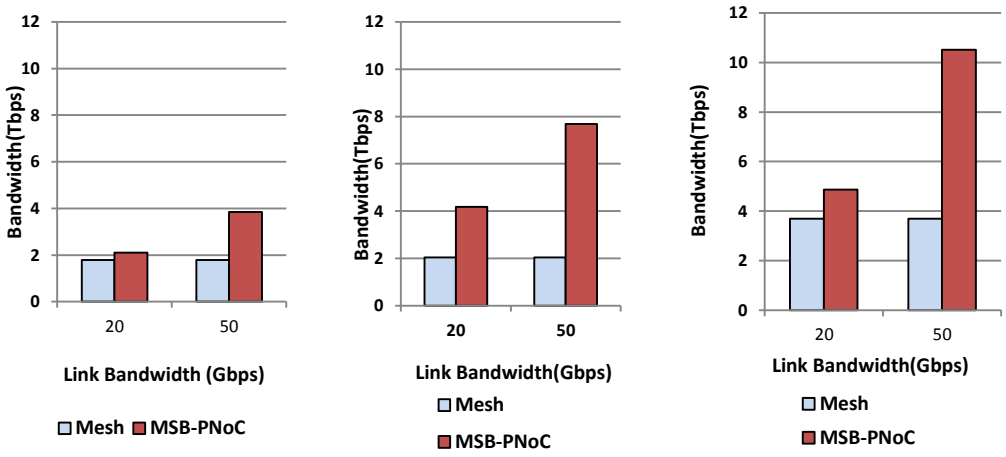


Figure 5-1 :16 cluster NoC BER Comparison

Because the MSB was designed specifically to decrease BER, for any given launched power, the MSB has the lowest BER, with the effect becoming more prominent for higher power values. The MSB design has a lower path length than Corona or 2DFT, resulting in a decrease in transmission errors and a lower BER. In general, a higher launched power leads to a stronger signal and more reliability in

general. As reported in [8], the highest feasible launched power per wavelength is 1.5mW. Any higher than that and the MRRs can experience resonance shifts. MRRs have a nonlinear mechanism known as free carrier dispersion (FCD), which can cause shifts in the resonant frequency of the MRRs at a faster rate than feedback loops are able to account for, causing unpredictable results. In the MSB, this maximum launched power value gives a worse-case BER of 10^{-14} for 20Gbps links, and 10^{-9} for 50Gbps [7]. It is assumed that as the bandwidth of the photonic links increases, the overall performance of the MSB-PNoC will also increase. Typical BER in data transfer over wireline links are of the order of 10^{-12} to 10^{-15} [14]. Hence, with 20Gbps photonic links the BER of an MSB architecture is comparable to that of an electronic mesh and not significantly worse.

Figure 5-2 shows how the peak sustainable bandwidth of the NoC is affected by the link bandwidth in a 64, 128, and 256 core system. The model uses non-blocking MSB architectures for better performance. The results show that as the bandwidth of the individual links increases, the overall data bandwidth also increases for any system size, since each individual link can support a higher data rate.



(a) **(b)** **(c)**

Figure 5-2: Data Bandwidth and BER of (a) 64, (b) 128, and (c) 256 core systems

The bandwidth of the mesh architecture with wireline links is shown for comparison. The bandwidth of photonic links does not have any effect on the system bandwidth of a mesh architecture, so the value is the same for both cases. In the 64 core system, the 20Gbps links only exhibited a slightly better bandwidth than the mesh. A substantial improvement was still present in the 64 core system with 50Gbps links. The true benefit of the MSB architecture becomes much more evident for the larger system sizes. The mesh architectures are not scalable, while MSB is designed for scalability, so as the system size increases, the advantage of MSB is much greater.

Table 5-1: Average and Maximum Path Length in Number of Hops, Mesh vs. MSB-PNoC

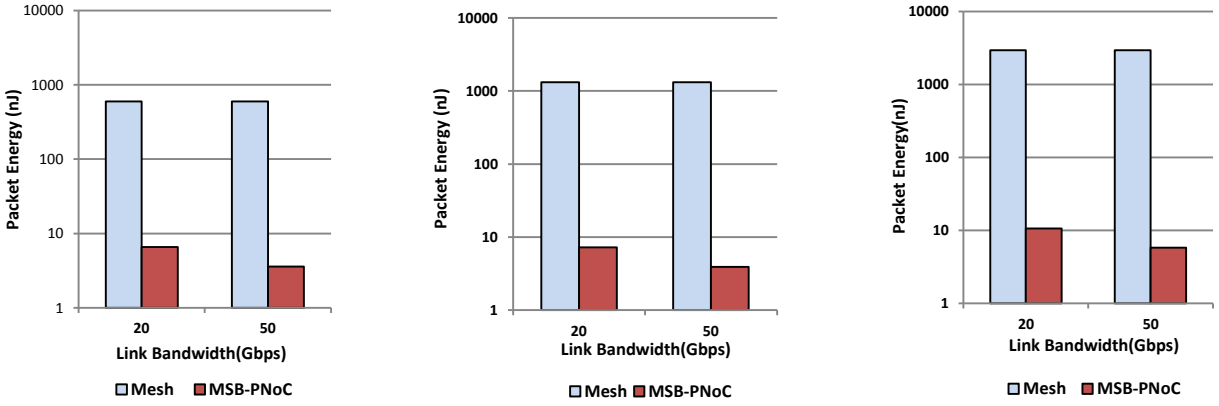
System Size	Mesh		MSB-PNoc	
	Avg	Max	Avg	Max
64	5.33	14	2.12	3
128	8	22	2.32	3
256	10.67	30	3.41	7

The cause of this is the difference in path lengths between distant cores. In a mesh network, the path lengths increase significantly as system size grows, but in an MSB system, the path length increases, but to a far lesser extent. Table 5-1 shows the maximum as well as the average path length in number of hops between cores in a mesh and MSB architecture. The average was computed using the equation

$$\bar{h} = \frac{\sum_{\forall ij} h_{ij}}{N(N-1)} \qquad (16)$$

where h_{ij} is the path length between cores i and j, measured in total number of hops. Because of this shorter path length, packets reach destinations quicker resulting in a much higher bandwidth gain for MSB-PNoC systems compared to conventional mesh networks, even with similar BERs, for large system sizes.

5.2. Packet Energy Dissipation



(a) (b) (c)

Figure 5-3: Packet Energy vs. Link Bandwidth for (a) 64, (b) 128, and (c) 256 Core Architecture

Figure 5-3 shows the average packet energy dissipation for all system sizes considered in this research. Again, both the conventional mesh and MSB were compared. Values for the energy dissipation of the modulators, demodulators, and routers for the MSB were obtained from [6]. Packet energy is considered to be the average energy dissipation to transfer packets from source to destination. The total packet energy dissipated for all packets was totaled, and divided by the total number of packets transferred. Since data is transferred through the low-power photonic waveguides, the energy dissipated by the MSB architecture is order of magnitude less than the conventional mesh. When the link bandwidth is increased, the system is able to transfer all of the flits faster, so the packet energy dissipation is decreased. This is seen in the figure as well, as the 50Gbps links for all system sizes exhibit a lower average energy dissipation. However, the lower energy dissipation

comes at the cost of reliability, because higher bandwidth in the links requires more channels in the waveguides, increasing adjacent channel crosstalk. As such, there is a trade-off between packet energy dissipation and reliability of the PNoC architecture. An example of this trade-off can be seen in Figure 5-4, 5-5, and 5-6. The packet energy improves significantly when the link bandwidth increases, but the system also shows an increase in BER, showing a decrease in reliability. For the larger system sizes, the energy benefit becomes less pronounced with the increase in individual link bandwidth. A possible reason for this effect is that the 256-core system size results in many more source-destination pairs needing several cycles to route compared to 64- or 128-core systems. The relative energy dissipated in modulation and demodulation to the IGBs is therefore higher in the 256-core system, meaning the relative energy dissipated across the links is lower. Because of this, increasing the link bandwidth improves the overall energy dissipation, but the improvement is relatively lower because the links account for a lower percentage of the overall energy.

5.3. Comparisons to 2DFT Photonic NoCs

The 2DFT architecture is one of the recent PNoC architectures proposed in the literature. A 2DFT system of 128 cores was also compared to the MSB-PNoC of the same size. This experiment took into account path multiplicity for the 2DFT architecture, as well as non-blocking for the MSB to analyze the best performance of each by including several parallel paths for each source/destination pair. Figure 5-7 shows that the MSB-PNoC has both a higher bandwidth and lower packet energy

than the 2DFT architecture. This could be due to the fact that in order to maintain full path multiplicity for larger system sizes, the 2DFT system requires a much more complex design, greatly increasing the number of MRRs, which in turn increases the energy dissipation as well as decreases reliability. The reliability-aware design of the MSB limits data transmission loss and crosstalk interference when compared to the 2DFT architecture. Both photonic NoCs have significantly lower packet energy dissipation than a conventional mesh, as well as a much higher sustainable bandwidth.

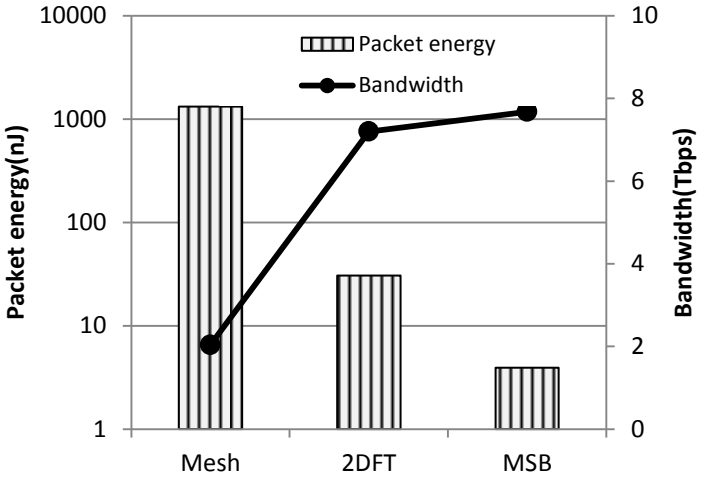


Figure 5-4: Packet Energy and Bandwidth of 128-Core NoCs

5.4. Performance Evaluation with Non-Uniform Traffic

The performance of the MSB-PNoC is also evaluated for synthetic and application-specific non-uniform traffic patterns. For synthetic traffic patterns, hotspot traffic and transpose traffic is selected to test the effect on MSB

performance. Hotspot traffic involves a single core being designated as the "hotspot", and all other cores sending 10% of their data to only that core. Transpose traffic has all cores only sending data to the diagonally opposite core in the network. For example, in a 64-core system, core number 1 would only send to core number 64 and vice versa, number 2 to number 59, and so on. For application-specific traffic, a 256-point Fast Fourier Transform (FFT) application is considered, with each core performing a 4-point radix-2 FFT computation. This model is used to calculate the source and destination cores that would be paired in a real-life, practical application. Figure 5-8 shows the bandwidth and packet energy dissipation for these traffic patterns under the same test conditions.

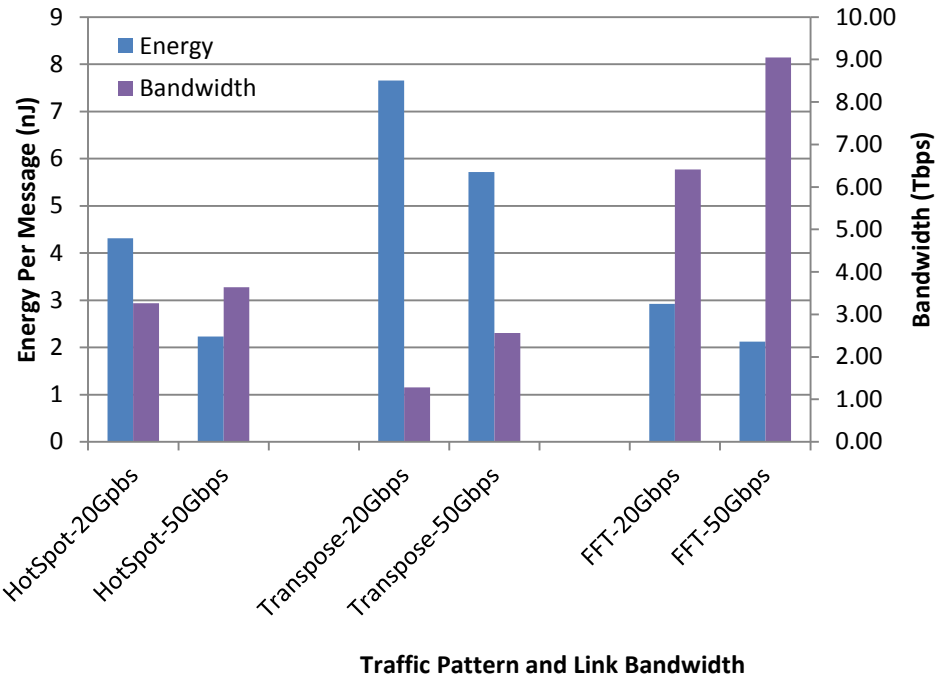


Figure 5-5: Packet Energy and Bandwidth of 128-Core MSB with Non-Uniform Traffic Patterns

One notable result from this experiment is that the transpose traffic pattern yields much higher number for energy per message than the other traffic types. This

is because the transpose pattern results in a large distance between many pairs of cores, because it uses diagonally opposite cores. This longer distance leads to longer data transfers, which would be expected to dissipate more energy. Hotspot yields lower energy results compared to transpose because the non-blocking architecture allows the system to avoid congestion around the hotspot core, causing the energy per message to remain lower than with transpose traffic. For the FFT pattern, the characteristic butterfly algorithm used in computation results in a particular pairing of cores, most of which result in shorter path length than the diametrically opposed pairing of the transpose traffic. This, in turn, results in faster transfers and lower energy dissipated. The other trends match those found in previous experiments, with the higher bandwidth links having higher overall bandwidth and lower energy dissipation.

5.5. Area Overhead

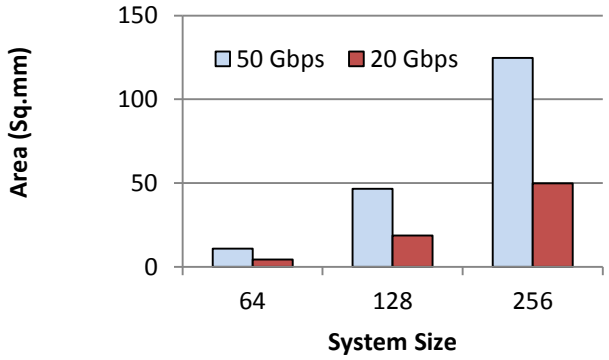


Figure 5-6: Area Overhead of the MSB-PNoC

The area overheads of the MSB are shown in figure 5-9. As the link bandwidth increases, the system needs a higher degree of WDM, which requires more photonic devices within the network. This leads to a larger area, increasing the

overhead. In order to ensure the architecture is non-blocking, the system needs to have parallel busses for concurrent communication between pairs of cores. This results in much greater performance, as there is no possibility of the links reaching a deadlock state, but also requires redundant hardware, further increasing the area overhead. With 20Gbps links the area overheads of the photonic components is around $50mm^2$ which is only 12.5% of the $400mm^2$ die area. However, with 50Gbps links this overhead increases to about 31.25%. This creates a trade-off between area and performance, in which performance can be sacrificed if the overall area were of a higher priority.

Chapter 6 Conclusions and Future Work

In this work a proposed photonic Network-on-Chip architecture with emphasis on being reliability-aware was designed and analyzed. Reliability analysis was taken into account, as well as experimental results were evaluated to compare the bandwidth, packet energy dissipation, and area overhead of the MSB-PNoC with other network architectures. This chapter summarizes the overall findings of this thesis work.

In comparing the bandwidth of the MSB-PNoC to that of a 2DFT PNoC, as well as a conventional mesh wireline network, the MSB architecture yielded slightly higher results for a 64-core system size, with the margin increasing for larger networks. Both the 2DFT and MSB outperformed the mesh network, with the MSB slightly improving on the 2DFT results as well. The MSB system improved on the mesh network bandwidth by nearly a factor of 4, and by close to 10% over 2DFT.

The low-power properties of photonic networks led to similar results when packet energy dissipation was analyzed. The 2DFT network showed large improvements over the conventional mesh network, because of the relatively high energy dissipation of wireline links. The MSB further improved on those values, exhibiting the lower dissipation of the networks tested.

A major advantage the MSB-PNoC has over mesh networks and some other PNoCs is its scalability. This fact was shown in that as the system size increases, the average and maximum path lengths from core to core increases at a much slower rate for the MSB PNoC when compared to a conventional mesh. This is a main factor

for why the MSB is able to expand the bandwidth and energy dissipation advantage it has over the other architectures, particularly for larger systems.

Non-uniform traffic patterns were analyzed to ensure the random traffic models were consistent with the behavior of the MSB PNoC under more specific circumstances. Hotspot, Transpose, and FFT traffic patterns were tested, with similar results to the uniform traffic model, lending itself to the fact that these are typical results.

Area overhead was also taken into account, especially due to the fact that scalability is a major advantage for the MSB. Higher system sizes necessitate the need for greater numbers of photonic devices, resulting in increasingly higher area overheads with larger system sizes. As sizes continue to increase, it becomes more important to consider the trade-off of area vs. performance.

The future challenges involved in improving the MSB design could include improvement on the base-level photonic devices. Since area overhead will continually increase as system sizes increase, creating devices that are smaller, or devices that exhibit lower levels of interference and crosstalk, would allow the system-size to increase without a sharp increase in area, or at least improve performance enough to make the area/performance trade-off more preferable. If scaling were to continue to 512- or 1028-core system sizes, this trade-off would become much more important. Additionally, analyzing the system under different system sizes, traffic patterns, etc. would provide further information for comparing with other photonic NoCs, which is more important since this testing showed definitively that PNoCs improve greatly over mesh architectures.

Bibliography

1. A. Shacham, K Bergman, L. Carloni, “Photonic Network-on-Chip for Future Generations of Chip Multi-Processors”, IEEE Transactions on Computers, Vol. 57, no. 9, 2008, pp. 1246-1260.
2. L. Benini and G. D. Micheli, “Networks on Chips: A New SoC Paradigm,” IEEE Computer, Vol. 35, Issue 1, January 2002, pp. 70-78.
3. W. Dally, B. Towles, “Route Packets, Not Wires: On-Chip Interconnection Networks," *Design Automation Conference, 2001. Proceedings* , vol., no., pp.684,689, 2001
4. B Feero, P. Pande, “Networks-on-Chip in a Three-Dimensional Environment: A Performance Evaluation," *Computers, IEEE Transactions on* , vol.58, no.1, pp.32,45, Jan. 2009
5. A. Ganguly, K. Chang, S. Deb, P. Pande, B. Belzer, C. Teuscher, “Scalable Hybrid Wireless Network-on-Chip Architectures for Multi-Core Systems s," *Computers, IEEE Transactions on* , vol.60, no.10, pp.1485,1502, Oct. 2011
6. L. P. Carloni, P. Pande and Y. Xie, "Networks-on-Chip in Emerging Interconnect Paradigms: Advantages and Challenges", Proceedings of the IEEE International Symposium on Networks-On-Chip, 10-13 May 2009.
7. I. Datta, D. Datta and P. P. Pande, “BER-based Power Budget Evaluation for Optical Interconnect Topologies in NoCs”, Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS 2012.

8. K. Preston, N. Sherwood-Droz, J. Levy, M. Lipson, "Performance Guidelines for WDM Interconnects Based on Silicon Microring Resonators," CLEO 2011 - Laser Applications to Photonic Applications, OSA Technical Digest (CD).
9. D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. Beausoleil, J. Ahn, "Corona: System Implications of Emerging Nanophotonic Technology," Proc. of IEEE International Symposium on Computer Architecture (ISCA), 21-25 June, 2008, pp. 153-164.
10. Joshi, A.; Batten, C.; Yong-Jin Kwon; Beamer, S.; Shamim, I.; Asanovic, K.; Stojanovic, V., "Silicon-photonic clos networks for global on-chip communication," *Networks-on-Chip, 2009. NoCS 2009. 3rd ACM/IEEE International Symposium on* , vol., no., pp.124,133, 10-13 May 2009
11. P. Pande, C. Grecu, M. Jones, A. Ivanov, R. Saleh, "Performance Evaluation and Design Trade-offs for Network-on-chip Interconnect Architectures", IEEE Transactions on Computers, Vol. 54, No. 8, August 2005, pp. 1025-1040.
12. A. Ganguly, Partha P. Pande and Benjamin Belzer, "Crosstalk-Aware Channel Coding Schemes for Energy Efficient and Reliable NoC Interconnects", IEEE Transactions on VLSI Vol. 17, No.11, November 2009, pp. 1626-1639.TVLSI
13. A. Ganguly, P. Wettin, K. Chang, P. Pande, "Complex Network Inspired Fault-Tolerant NoC Architectures with Wireless Links," *Networks on Chip (NoCS), 2011 Fifth IEEE/ACM International Symposium on,* pp. 169,176, 1-4 May 2011

14. S. R. Sridhara, and N. R. Shanbhag, “Coding for System-on-Chip Networks: A Unified Framework”, IEEE Transactions on Very Large Scale Integration (TVLSI) Systems, vol. 13, no. 6, June 2005,pp.655-667